Claims

WHAT IS CLAIMED IS:

1. A method comprising:

computing a first probe estimate from a first probe sequence transmitted over a network from a first device to a second device;

computing a second probe estimate from a second probe sequence transmitted over the network from the first device to the second device, the first probe sequence providing a different load to the network than the second probe sequence; and

estimating a sustainable capacity of the network based on the first probe estimate and the second probe estimate.

2. The method of claim 1 wherein the sustainable capacity is further based on a minimum delay time of individual round-trip delay times associated with the first probe sequence and the second probe sequence.

3. The method of claim 1 wherein the sustainable capacity of the network relates to data traffic flowing from the first device to the second device, and further comprising:

computing a third probe estimate from a third probe sequence transmitted over the network from the second device to the first device;

computing a fourth probe estimate from a fourth probe sequence transmitted over the network from the second device to the first device, the third probe sequence providing a different load to the network than the fourth probe sequence; and

estimating a reverse sustainable capacity of the network for data traffic from the second device and the first device based on the third probe estimate and the fourth probe estimate.

4. The method of claim 1 wherein the first probe sequence includes probe packets periodically transmitted from the first device to the second device.

5. The method of claim 1 wherein the estimating operation comprises computing an algorithm substantially of the form:

$$SusCap_{est} = \delta m \frac{\hat{D}_2 - D_{min}}{\hat{D}_2 - \hat{D}_1}$$

wherein $SusCap_{est}$ represents the sustainable capacity value, δm represents a load difference between the second probe sequence and the first probe sequence, $\hat{D}_2$ represents the second probe estimate, $\hat{D}_1$ represents the first probe estimate, and D_{min} represents a minimum delay time of individual round-trip delay times associated with the first load probe sequence and the second load probe sequence.

6. The method of claim 1 wherein the estimating operation comprises computing an algorithm substantially of the form:

$$SusCap_{est} = \delta m \frac{\hat{D}_2}{\hat{D}_2 - \hat{D}_1}$$

wherein $SusCap_{est}$ represents the sustainable capacity value, δm represents a load difference between the second probe sequence and the first probe sequence, $\hat{D}_2$ represents the second probe estimate, and $\hat{D}_1$ represents the first probe estimate.

7. The method of claim 1 wherein neither the first probe sequence nor the second probe sequence saturate the network.

8. The method of claim 1 wherein the second probe sequence loads the network more than the first probe sequence.

9. The method of claim 1 further comprising:

generating a first probe sequence, wherein each probe packet in the first probe sequence has a unique signature.

10. The method of claim 1 further comprising:

timestamping each packet in the first probe sequence prior to transmission to the second network device.

11. The method of claim 1 further comprising:

timestamping each response to each probe packet in the first probe sequence after to reception of the probe packet from the second network device.

12. The method of claim 1 wherein each probe estimate is an average round-trip delay time.

13. The method of claim 1 wherein each probe estimate is a median round-trip delay time.

14. The method of claim 1 wherein each probe estimate is a range of round-trip delay times.

15. The method of claim 1 wherein each probe estimate is a standard deviation of round-trip delay times.

16. A computer program product encoding a computer program for executing on a computer system a computer process, the computer process comprising:

computing a first probe estimate from a first probe sequence transmitted over a network from a first device to a second device;

computing a second probe estimate from a second probe sequence transmitted over the network from the first device to the second device, the first probe sequence providing a different load to the network than the second probe sequence; and

estimating a sustainable capacity of the network based on the first probe estimate and the second probe estimate.

17. The computer program product of claim 16 wherein the sustainable capacity value is further based on a minimum delay time of individual round-trip delay times associated with the first probe sequence and the second probe sequence.

18. The computer program product of claim 16 wherein the sustainable capacity of the network relates to data traffic flowing from the first device to the second device, and further comprising:

computing a third probe estimate from a third probe sequence transmitted over the network from the second device to the first device;

computing a fourth probe estimate from a fourth probe sequence transmitted over the network from the second device to the first device, the third

probe sequence providing a different load to the network than the fourth probe sequence; and

estimating a reverse sustainable capacity value in the network characterizing sustainable capacity of the network for data traffic from the second device and the first device based on the third probe estimate and the fourth probe estimate.

19. The computer program product of claim 16 wherein the first probe sequence includes probe packets periodically transmitted from the first device to the second device.

20. The computer program product of claim 16 wherein the estimating operation comprises computing an algorithm substantially of the form:

$$SusCap_{est} = \delta m \frac{\hat{D}_2 - D_{\min}}{\hat{D}_2 - \hat{D}_1}$$

wherein $SusCap_{est}$ represents the sustainable capacity value, δm represents a load difference between the second probe sequence and the first probe sequence, $\hat{D}_2$ represents the second probe estimate, $\hat{D}_1$ represents the first probe estimate, and $D_{\min}$ represents a minimum delay time of individual round-trip delay times associated with the first load probe sequence and the second load probe sequence.

21. The computer program product of claim 16 wherein the estimating operation comprises computing an algorithm substantially of the form:

$$SusCap_{est} = \delta m \frac{\hat{D}_2}{\hat{D}_2 - \hat{D}_1}$$

wherein $SusCap_{est}$ represents the sustainable capacity value, δm represents a load difference between the second probe sequence and the first probe sequence, $\hat{D}_2$ represents the second probe estimate, and $\hat{D}_1$ represents the first probe estimate.

22. The computer program product of claim 16 wherein neither the first probe sequence nor the second probe sequence saturate the network.

23. The computer program product of claim 16 wherein the second probe sequence loads the network more than the first probe sequence.

24. The computer program product of claim 16 wherein the computer process further comprises:

generating a first probe sequence, wherein each probe packet in the first probe sequence has a unique signature.

25. The computer program product of claim 16 wherein the computer process further comprises:

timestamping each packet in the first probe sequence prior to transmission to the second network device.

26. The computer program product of claim 16 wherein the computer process further comprises:

timestamping each response to each probe packet in the first probe sequence after to reception of the probe packet from the second network device.

27. The computer program product of claim 16 wherein each probe estimate is an average round-trip delay time.

28. The computer program product of claim 16 wherein each probe estimate is a median round-trip delay time.

29. The computer program product of claim 16 wherein each probe estimate is a range of round-trip delay times.

30. The computer program product of claim 16 wherein each probe estimate is a standard deviation of round-trip delay times.

31. A system comprising:

a probe sequence generator that generates a first probe sequence and a second probe sequence;

a communication module that transmits the first probe sequence and the second probe sequence to a target, receives a first response sequence associated with the first probe sequence and a second response sequence associated with the second probe sequence; and computes a first probe estimate from the first probe sequence and a second probe estimate from the second probe sequence; and

a probe performance analyzer that estimates a sustainable capacity in the network characterizing sustainable capacity of the network for data traffic from the first device and the second device based on the first probe estimate and the second probe estimate.

32. The system of claim 31 wherein the sustainable capacity is further based on a minimum delay time of individual round-trip delay times associated with the first probe sequence and the second probe sequence.

33. The system of claim 31 wherein the first probe sequence includes probe packets periodically transmitted from the first device to the second device.

34. The system of claim 31 wherein the probe performance analyzer computes an algorithm substantially of the form:

$$SusCap_{est} = \delta m \frac{\hat{D}_2 - D_{min}}{\hat{D}_2 - \hat{D}_1}$$

wherein $SusCap_{est}$ represents the sustainable capacity value, δm represents a load difference between the second probe sequence and the first probe sequence, $\hat{D}_2$ represents the second probe estimate, $\hat{D}_1$ represents the first probe estimate, and $D_{\min}$ represents a minimum delay time of individual round-trip delay times associated with the first load probe sequence and the second load probe sequence.

35. The system of claim 31 wherein the probe performance analyzer computes an algorithm substantially of the form:

$$SusCap_{est} = \delta m \frac{\hat{D}_2}{\hat{D}_2 - \hat{D}_1}$$

wherein $SusCap_{est}$ represents the sustainable capacity value, δm represents a load difference between the second probe sequence and the first probe sequence, $\hat{D}_2$ represents the second probe estimate, and $\hat{D}_1$ represents the first probe estimate.

36. The system of claim 31 wherein neither the first probe sequence nor the second probe sequence saturate the network.

37. The system of claim 31 wherein each probe estimate is an average round-trip delay time.

38. The system of claim 31 wherein each probe estimate is a median round-trip delay time.

39. The system of claim 31 wherein each probe estimate is a range of round-trip delay times.

40. The system of claim 31 wherein each probe estimate is a standard deviation of round-trip delay times.

41. A method comprising:

computing a first round-trip delay trend associated with a first probe sequence transmitted over a network from a first device to a second device;

computing a second round-trip delay trend associated with a second probe sequence transmitted over the network from the first device to the second device;

estimating a maximum capacity and a background load of the network based on the first round-trip delay trend and the second round-trip delay trend; and

estimating a sustainable capacity of the network based on the maximum capacity and the background load of the network.

42. The method of claim 41 wherein a number of probe packets in the first load probe sequence equals a number of probe packets in the second load probe sequence.

43. The method of claim 41 wherein a size of all probe packets in the first load probe sequence equals a size of all probe packets in the second load probe sequence and an inter-probe gap between probe packets in the first load probe sequence does not equal an inter-probe gap between probe packets in the second load probe sequence.

44. The method of claim 41 wherein an inter-probe gap between probe packets in the first load probe sequence equals an inter-probe gap between probe packets in the second load probe sequence and a size of all probe packets in the first load probe sequence does not equal a size of all probe packets in the second load probe sequence

45. The method of claim 41 wherein the operation of computing the first round-trip delay trend comprises:

computing the average round-trip delay difference associated with the first probe sequence.

46. A computer program product encoding a computer program for executing on a computer system a computer process, the computer process comprising:

computing a first round-trip delay trend associated with a first probe sequence transmitted over a network from a first device to a second device;

computing a second round-trip delay trend associated with a second probe sequence transmitted over the network from the first device to the second device;

estimating a maximum capacity and a background load of the network based on the first round-trip delay trend and the second round-trip delay trend; and

estimating a sustainable capacity of the network based on the maximum capacity and the background load of the network.

47. The computer program product of claim 46 wherein a number of probe packets in the first load probe sequence equals a number of probe packets in the second load probe sequence.

48. The computer program product of claim 46 wherein a size of all probe packets in the first load probe sequence equals a size of all probe packets in the second load probe sequence and an inter-probe gap between probe packets in the first load probe sequence does not equal an inter-probe gap between probe packets in the second load probe sequence.

49. The computer program product of claim 46 wherein an inter-probe gap between probe packets in the first load probe sequence equals an inter-probe gap between probe packets in the second load probe sequence and a size of all probe

packets in the first load probe sequence does not equal a size of all probe packets in the second load probe sequence

50. The computer program product of claim 46 wherein the operation of computing the first round-trip delay trend comprises:

computing the average round-trip delay difference associated with the first probe sequence.

51. A system comprising:

a probe sequence generator that generates a first probe sequence and a second probe sequence;

a communication module that transmits the first probe sequence and the second probe sequence to a target, receives a first response sequence associated with the first probe sequence and a second response sequence associated with the second probe sequence; and computes first round-trip delay times of the first probe sequence and second round-trip delay times of the second probe sequence;

a probe performance analyzer that computes a first round-trip delay trend associated with a first probe sequence and a second round-trip delay trend associated with a second probe sequence, estimates a maximum capacity and a background load of the network based on the first round-trip delay trend and the second round-trip delay trend, and estimates a sustainable capacity of the network based on the maximum capacity and the background load of the network.

52. The system of claim 51 wherein a number of probe packets in the first load probe sequence equals a number of probe packets in the second load probe sequence.

53. The system of claim 51 wherein a size of all probe packets in the first load probe sequence equals a size of all probe packets in the second load probe sequence and an inter-probe gap between probe packets in the first load probe sequence does not equal an inter-probe gap between probe packets in the second load probe sequence.

54. The system of claim 51 wherein an inter-probe gap between probe packets in the first load probe sequence equals an inter-probe gap between probe packets in the second load probe sequence and a size of all probe packets in the first load probe sequence does not equal a size of all probe packets in the second load probe sequence

55. The system of claim 51 wherein the probe performance analyzer further computes the average round-trip delay difference associated with the first probe sequence.

56. A method comprising:

determining a network saturation window by incrementally increasing the load of probe packet sequence between a client and a server on a network until at least one probe packet is lost;

detecting a time period associated with the network saturation window; and

estimating a sustainable capacity of the network between the client and the server from a ratio of the network saturation window to the time period.

57. The method of claim 56 wherein detecting the time period comprises:

detecting a minimum round-trip delay time on the network between the client and the server, wherein the minimum round-trip delay time constitutes the time period.

58. The method of claim 56 wherein detecting the time period comprises sustaining traffic within the network saturation window on the network between the client and the server for a predetermined period of time and estimating the sustainable capacity of the network comprises computing the number of bytes per unit time transmitted in the network saturation window during the time period.

59. A computer program product encoding a computer program for executing on a computer system a computer process, the computer process comprising:

determining a network saturation window by incrementally increasing the load of probe packet sequence between a client and a server on a network until at least one probe packet is lost;

detecting a time period associated with the network saturation window; and

estimating a sustainable capacity of the network between the client and the server from a ratio of the network saturation window to the time period.

60. The computer program product of claim 59 wherein detecting the time period comprises:

detecting a minimum round-trip delay time on the network between the client and the server, wherein the minimum round-trip delay time constitutes the time period.

61. The computer program product of claim 59 wherein detecting the time period comprises sustaining traffic within the network saturation window on the network between the client and the server for a predetermined period of time and estimating the sustainable capacity of the network comprises computing the number of bytes per unit time transmitted in the network saturation window during the time period.

62. A system comprising:

a communications module that transmits a probe packet sequence of incrementally increasing load between a client and a server on a network; and

a sustainable capacity probe module that determines a network saturation window from the probe packet sequence based on when at least one probe packet is lost and estimates a sustainable capacity of the network between the client and the server from a ratio of the network saturation window to a determined time period.

63. The system of claim 62 wherein the determined time period is a minimum round-trip delay time detected between the receiver that the sender.

64. The system of claim 62 wherein the determined time period is a period of transmitting sustained traffic within the network saturation window on the network between the client and the server.